

SỬ DỤNG TRÍ TUỆ NHÂN TẠO CÓ ĐẠO ĐỨC VÀ TRÁCH NHIỆM: ĐỀ XUẤT KHUNG HÀNH ĐỘNG 3T CHO SINH VIÊN VIỆT NAM

USING ETHICAL AND RESPONSIBLE ARTIFICIAL INTELLIGENCE: PROPOSING THE 3T ACTION FRAMEWORK FOR VIETNAMESE STUDENTS

Nguyễn Duy Hải^{1,+},
Tô Hồng Nam²

¹Khoa Công nghệ số liên ngành - Đại học Phenikaa;

²Cục Khoa học, Công nghệ và Thông tin - Bộ Giáo dục và Đào tạo

+Tác giả liên hệ • Email: hai.nguyenduy@phenikaa-uni.edu.vn

Article history

Received: 30/3/2026

Accepted: 29/4/2026

Published: 05/6/2026

Keywords

AI ethics, academic integrity,
AI literacy, prompt
engineering, responsible AI,
higher education, generative
AI

ABSTRACT

Generative Artificial Intelligence (Gen AI) is becoming increasingly prevalent in Vietnamese higher education, creating both learning opportunities and challenges related to ethics and responsibility. The AI Ready ASEAN Research (2026) indicates that 89.3% of students in Vietnam have used generative AI, yet only about one quarter have received formal training. This study employs an interpretive qualitative design, with systematic classroom observation as the primary data collection technique. Data were collected from six classes across six digital technology programs at a private Vietnamese university during the 2025-2026 academic year, interpreted alongside prior studies conducted in Vietnam. Three recurring behavioural gaps are identified: (1) ineffective AI use due to limited prompt-writing skills; (2) submission of unverified AI-generated outputs; and (3) blurred awareness of ethical boundaries in AI use. In response, this paper proposes the “3T” framework (named after the Vietnamese concepts of Thành thạo (Mastery), Trung thực (Integrity), and Trách nhiệm (Responsibility)) as a practical guide for students and offers policy recommendations aligned with Vietnam’s current higher education and digital transformation agenda.

1. Mở đầu

Chỉ trong vòng hai năm kể từ khi ChatGPT ra mắt vào cuối năm 2022, trí tuệ nhân tạo (AI) tạo sinh đã tác động sâu sắc đến học tập và nghiên cứu ở bậc đại học. Sinh viên (SV) thường xuyên sử dụng AI để làm bài luận, lập trình, dịch thuật và chuẩn bị bài tập nhóm - thường là trước khi các cơ sở giáo dục đại học ban hành các quy định sử dụng AI trong học tập và nghiên cứu học thuật. Tại Việt Nam, mức độ phổ cập AI trong SV đã đạt ngưỡng rất cao: nghiên cứu AI Ready ASEAN do ASEAN Foundation và Google.org (2026) công bố cho thấy 89,3% SV Việt Nam đã sử dụng AI tạo sinh - cao hơn mức trung bình khu vực ASEAN. Tuy nhiên, nghịch lý đáng lo ngại là chỉ khoảng một phần tư SV được đào tạo chính thức về AI, trong khi gần hai phần ba giảng viên đã hoàn thành các khóa học liên quan đến AI. Khoảng cách này đặt ra câu hỏi cấp thiết: ranh giới giữa AI hỗ trợ học tập và vi phạm học thuật nằm ở đâu và làm thế nào để SV sử dụng AI một cách có đạo đức và trách nhiệm?

Các nghiên cứu về AI trong giáo dục đại học có thể được nhìn theo ba hướng chính. Hướng thứ nhất, tập trung vào thực trạng sử dụng AI của SV và cho thấy AI đã được tích hợp sâu vào hoạt động học tập, chủ yếu cho tìm kiếm thông tin và dịch thuật với nguy cơ phụ thuộc và mất kỹ năng tự học (Đặng Văn Em và cộng sự, 2024; Nguyễn Thị Lê Quỳnh và cộng sự, 2024; Nguyễn Thị Phương Thảo và cộng sự, 2024). Hướng thứ hai, nhấn mạnh đến liên chính học thuật trong thời đại AI và chỉ ra rằng SV tồn tại trong một “vùng xám” nhận thức - họ biết nộp bài do AI viết mà không kiểm tra lại là vi phạm, nhưng lại không rõ ranh giới giữa được phép và không được phép ở đâu (Chan, 2023, 2024; Perkins, 2023; Balalle và Pannilage, 2025). Hướng thứ ba, tập trung vào khung đạo đức và năng lực AI, nhấn mạnh rằng năng lực AI phải vượt ra ngoài kỹ năng kỹ thuật để bao gồm đạo đức, minh bạch và phù hợp văn hóa xã hội (UNESCO, 2021, 2024; Allen và Kendeou, 2024; Nguyen và cộng sự, 2023; Cui và cộng sự, 2026). Điều này cho thấy khoảng trống là: hiện còn thiếu các nghiên cứu định tính có hệ thống trong bối cảnh giáo dục đại học Việt Nam nhằm nhận diện các khoảng trống hành vi khi SV sử dụng AI và đề xuất một khung hành động bản địa hóa phù hợp với môi trường học thuật trong nước.

Về nền tảng lý thuyết, nghiên cứu dựa trên ba trụ cột. Thứ nhất, Khuyến nghị về đạo đức AI của UNESCO (2021) xác định ba nguyên tắc cốt lõi trong giáo dục: minh bạch (người dùng có quyền biết khi nào AI hỗ trợ tạo ra nội dung), trách nhiệm giải trình (AI không thể thay thế trách nhiệm cuối cùng của con người) và giám sát của con người (quyền kiểm soát và can thiệp luôn thuộc về con người). Thứ hai, các nghiên cứu về liên chính học thuật trong thời đại AI (Perkins, 2023; Chan, 2024; Balalle và Pannilage, 2025) xác định một phổ ba cấp độ vi phạm từ hỗ trợ được chấp nhận đến gian lận rõ ràng, với phần lớn SV tồn tại trong vùng xám ở giữa. Thứ ba, Khung ED-AI Lit của Allen và Kendeou (2024) và Khung năng lực AI của UNESCO (2024) cung cấp sáu thành phần năng lực AI trong giáo dục, trong đó Quyền tự chủ và Đạo đức là hai thành phần nền tảng nhất. Logic hình thành Khung “3T” xuất phát trực tiếp từ sự đối chiếu giữa ba khoảng trống hành vi quan sát được và ba trụ cột lý thuyết này: khoảng trống kỹ năng → cần “Thành thạo”; khoảng trống liên chính học thuật → cần “Trung thực”; khoảng trống ý thức đạo đức và trách nhiệm dài hạn → cần “Trách nhiệm”.

Trên cơ sở đó, nghiên cứu này tập trung trả lời ba câu hỏi: (1) Những khoảng trống hành vi phổ biến nào xuất hiện khi SV đại học Việt Nam sử dụng AI trong học tập?; (2) Các khoảng trống này phản ánh những vấn đề gì về đạo đức AI và liên chính học thuật?; (3) Khung hành động nào phù hợp để hướng dẫn SV Việt Nam sử dụng AI có đạo đức và trách nhiệm? Nghiên cứu kỳ vọng đóng góp gồm: (1) cung cấp bằng chứng hội tụ về thực trạng sử dụng AI từ quan sát lớp học định tính và các nghiên cứu trước tại Việt Nam; (2) đặt thực trạng sử dụng AI của SV Việt Nam trong bối cảnh các khung lý thuyết quốc tế; (3) đề xuất Khung “3T” (Thành thạo - Trung thực - Trách nhiệm) như công cụ thực hành được bản địa hóa cho SV Việt Nam, phù hợp với tinh thần Nghị quyết số 71-NQ/TW của Ban Chấp hành Trung ương (2025).

2. Phương pháp nghiên cứu

Nghiên cứu sử dụng thiết kế định tính diễn giải (interpretive qualitative design), trong đó quan sát có hệ thống của giảng viên (systematic instructor observation) là kỹ thuật thu thập dữ liệu chính; cách tiếp cận nghiên cứu hành động trong lớp học (classroom-based action research) được sử dụng để hỗ trợ quá trình phân tư và hình thành khung hành động ban đầu. Thiết kế này được lựa chọn vì mục tiêu nghiên cứu là khám phá và mô tả các mẫu hành vi trong bối cảnh tự nhiên của lớp học, không nhằm kiểm chứng giả thuyết hay đo lường quan hệ nhân quả. Khi chưa có công cụ khảo sát được kiểm chứng về hành vi sử dụng AI có đạo đức trong giáo dục đại học Việt Nam, quan sát thực địa có cấu trúc trở thành nguồn dữ liệu có độ sinh thái cao và phù hợp cho việc xây dựng lý thuyết.

Nghiên cứu được thực hiện trong đầu năm học 2025-2026 tại một trường đại học ngoài công lập Việt Nam. Dữ liệu được thu thập tại sáu lớp học (mỗi ngành một lớp), thuộc sáu chương trình đào tạo liên quan đến công nghệ số: Khoa học máy tính, Công nghệ thông tin, Kinh tế số, Thương mại điện tử, Truyền thông đa phương tiện và Logistics số, với tổng số khoảng 1.600 SV tham gia các phiên học. Mỗi chương trình đào tạo đại diện cho một nhóm đặc thù về mức độ sử dụng AI và loại nhiệm vụ học tập.

Dữ liệu sơ cấp được thu thập qua hai kênh có cấu trúc. Kênh thứ nhất là khung KWL (Know - Wonder - Learned) được triển khai dưới dạng các phiên tập huấn tương tác ngắn (mini workshop) tại mỗi lớp, sử dụng nền tảng Padlet để SV đăng tải ý kiến theo thời gian thực trên bảng tương tác ba cột K - W - L. Phân tích tập trung vào hai cột K (nhận thức hiện có) và W (nhu cầu tìm hiểu), nhằm phản ánh trực tiếp cách SV hiểu và đặt vấn đề về việc sử dụng AI trong học tập. Hoạt động KWL được dẫn dắt bởi một tập hợp câu hỏi định hướng nhất quán qua các lớp, đảm bảo so sánh liên lớp. Sáu câu hỏi xuất hiện với tần suất cao nhất trong cột W - tức là được nhiều SV ở nhiều lớp khác nhau đặt ra - được dùng làm đơn vị phân tích chính: (1) Dùng AI thế nào thì không bị coi là gian lận? (2) Có nên ghi rõ khi dùng AI không? (3) AI viết hộ bài thì sai ở đâu? (4) Dùng AI làm đồ án/code có bị coi là đạo văn không? (5) Đưa dữ liệu vào AI có nguy hiểm không? (6) Sau này đi làm, dùng AI có bị cấm không?

Kênh thứ hai là các mẫu hành vi được ghi nhận trong các buổi tư vấn bài tập, thảo luận dự án và phản hồi bài nộp. Ghi chép được thực hiện vào nhật ký thực địa ngay sau mỗi buổi để đảm bảo tính nhất quán. Chỉ các mẫu quan sát được trên nhiều lớp và ngành mới được giữ lại cho phân tích.

Về phân tích dữ liệu, dữ liệu từ Padlet được lưu trữ và xử lý thông qua phương pháp mã hóa chủ đề (thematic coding). Cụ thể, các câu hỏi và phản hồi của SV trong cột W được phân nhóm theo các mẫu ý nghĩa lặp lại (recurring meaning patterns) giữa các lớp. Tiêu chí xác định “tần suất cao” được định nghĩa là chủ đề xuất hiện nhất quán trong ít nhất bốn trong sáu lớp học. Dữ liệu được mã hóa theo từng chủ đề lớn (ví dụ: ranh giới đạo đức, kỹ năng viết lệnh, rủi ro dữ liệu), sau đó được tổng hợp thành các khoảng trống hành vi. Bối cảnh thu thập dữ liệu được mô tả tại bảng 1.

Để tăng cường độ tin cậy, quan sát sơ cấp được đối chiếu tam giác ở cấp độ chủ đề (thematic triangulation) với kết quả từ ba nghiên cứu độc lập tại các trường đại học Việt Nam (Đặng Văn Em và cộng sự, 2024; Nguyễn Thị Lê Quỳnh và cộng sự, 2024; Nguyễn Thị Phương Thảo và cộng sự, 2024) và dữ liệu khảo sát khu vực ASEAN (ASEAN Foundation và Google.org, 2026). Vai trò của đối chiếu là xác nhận mức độ hội tụ giữa phát hiện của nghiên cứu này và các nghiên cứu độc lập - không phải thay thế bằng chứng thực địa của bài. Sự hội tụ giữa các nguồn độc lập củng cố độ tin cậy mà không yêu cầu khái quát hóa thống kê.

Toàn bộ dữ liệu được ẩn danh trước khi phân tích và chỉ được sử dụng cho mục đích nghiên cứu học thuật. Tuy nhiên, phương pháp này có giới hạn: dữ liệu quan sát định tính không cho phép khái quát hóa thống kê và hơn 1.600 SV là quy mô bối cảnh lớp học - không phải 1.600 trường hợp được phân tích theo nghĩa định tính đầy đủ; đơn vị phân tích thực tế là các chủ đề hành vi lặp lại qua sáu lớp học.

Bảng 1. Mô tả bối cảnh thu thập dữ liệu

Chương trình đào tạo	Số lớp quan sát	Công cụ thu thập	Đơn vị phân tích
Khoa học máy tính/Công nghệ thông tin	2 lớp	KWL qua Padlet	Câu hỏi cột W; mẫu hành vi bài nộp
Kinh tế số/Thương mại điện tử	2 lớp	KWL qua Padlet	Câu hỏi cột W; mẫu hành vi bài nộp
Truyền thông đa phương tiện/Logistics số	2 lớp	KWL qua Padlet	Câu hỏi cột W; mẫu hành vi bài nộp
Tổng	6 lớp (~1.600 SV)	Padlet + nhật kí thực địa	6 chủ đề/câu hỏi tái xuất hiện; 3 mẫu hành vi nhất quán

(Nguồn: Tổng hợp của tác giả)

3. Kết quả nghiên cứu

3.1. Ba khoảng trống hành vi trong sử dụng AI của sinh viên Việt Nam

Ba khoảng trống hành vi phổ biến và nhất quán của SV khi sử dụng AI được xác định qua quan sát. Đáng chú ý, cả ba đều có sự hội tụ với kết quả từ các nghiên cứu độc lập tại các trường đại học Việt Nam khác, gợi mở xu hướng có khả năng mang tính rộng hơn một cơ sở giáo dục đại học đơn lẻ, dù cần kiểm chứng ở quy mô rộng hơn. Tổng hợp 3 khoảng trống này được mô tả ở bảng 2.

3.1.1. Khoảng trống 1: Sử dụng AI chưa hiệu quả do thiếu kỹ năng viết prompt

Phần lớn SV tương tác với AI qua câu lệnh đơn giản, thiếu ngữ cảnh và không có yêu cầu đầu ra cụ thể. Ví dụ điển hình quan sát được: thay vì viết “*Bạn là nhà phân tích dữ liệu. Hãy giải thích hồi quy tuyến tính cho SV Kinh tế số năm thứ hai bằng ví dụ dự báo doanh số bán lẻ, trong 300 từ, tránh thuật ngữ thống kê phức tạp*”, SV chỉ gõ “*giải thích hồi quy tuyến tính*”. Kết quả AI tạo ra mang tính chung chung và không đáp ứng yêu cầu cụ thể của nhiệm vụ.

Ngoài ví dụ trên, trong cột W của Padlet, nhiều SV đặt các câu hỏi cho thấy họ chưa biết cách định hướng AI: “*Làm sao để AI cho ra bài viết đúng ý mình?*”, “*Tại sao AI hiểu sai yêu cầu của mình?*” Phát hiện này nhất quán với Đặng Văn Em và cộng sự (2024), Phung và cộng sự (2025): SV Việt Nam chủ yếu dùng AI cho tìm kiếm thông tin và dịch thuật đơn giản, ít áp dụng cho phân tích hay tư duy phản biện. Nguyên nhân gốc rễ là thiếu tập huấn kỹ năng viết prompt trong chương trình đào tạo.

3.1.2. Khoảng trống 2: Nộp kết quả AI chưa qua kiểm chứng

Xu hướng phổ biến thứ hai là SV sao chép kết quả AI và nộp bài mà không đọc lại, không đối chiếu với nguồn độc lập, không điều chỉnh theo yêu cầu cụ thể. Hành vi này khiến SV rơi vào hiện tượng ảo giác AI (hallucination) - xu hướng của mô hình ngôn ngữ lớn tạo ra thông tin sai nhưng có vẻ đáng tin cậy. Balalle và Pannilage (2025) xác định đây là một trong những rủi ro học tập nghiêm trọng nhất.

Một tình huống điển hình ghi nhận được: trong buổi phản hồi bài nộp môn Kinh tế số, SV nộp bài phân tích có trích dẫn một nghiên cứu không tồn tại do AI tạo ra - khi được hỏi, SV thừa nhận đã copy trực tiếp mà không kiểm tra lại. Nguyễn Thị Phương Thảo và cộng sự (2024) phát hiện dù hơn 70% trong số 1.101 SV tại Huế cho rằng “*chỉ nên dùng ChatGPT như công cụ hỗ trợ*”, hành vi thực tế lại mâu thuẫn với nhận thức đó: áp lực thời gian và niềm tin “*AI luôn đúng*” dẫn đến bài nộp không được kiểm chứng.

3.1.3. Khoảng trống 3: Nhận thức mờ nhạt về ranh giới đạo đức

Khoảng trống thứ ba và có hệ quả lâu dài nhất là SV thiếu rõ ràng về ranh giới giữa sử dụng AI được phép và không được phép. Sáu câu hỏi thu được qua Padlet từ hơn 1.600 SV không phải là câu hỏi của người có ý gian lận - đây là câu hỏi của người thực sự không biết ranh giới ở đâu. Ví dụ, câu hỏi “*AI viết hộ bài thì sai ở đâu?*” xuất

hiện ở cả sáu lớp và nhận được nhiều phản hồi đồng tình từ các SV khác, cho thấy đây là điểm mù nhận thức phổ biến chứ không phải cá biệt.

Phát hiện này hội tụ với Chan (2023): SV tồn tại trong “vùng xám” nhận thức và đây là nguồn gốc của khái niệm “AI-giarism” - đạo văn từ AI. Tại King’s Business School, 74% SV không công bố có sử dụng AI dù có chính sách bắt buộc - chủ yếu do hiểu lầm hơn là vi phạm cố ý (Perkins, 2023). Tại Việt Nam, UNESCO (2025) xác nhận chưa có cơ chế thể chế nào cung cấp hướng dẫn rõ ràng về sử dụng AI có đạo đức trong môi trường học thuật - một khoảng trống chính sách cần được khẩn cấp bổ sung.

Bảng 2. Tổng hợp ba khoảng trống về sử dụng AI trong học tập tại Việt Nam

Khoảng trống	Biểu hiện	Hậu quả	Bằng chứng thực nghiệm tại Việt Nam
1. Đặt lệnh kém hiệu quả (KT1)	Câu lệnh đơn giản, thiếu ngữ cảnh, không có yêu cầu đầu ra cụ thể	Kết quả AI chất lượng thấp; không khai thác được tiềm năng của AI	89,2% SV Đại học Quốc gia TP. Hồ Chí Minh dùng gói miễn phí với câu lệnh cơ bản (Đặng Văn Em và cộng sự, 2024)
2. Nộp bài AI chưa kiểm chứng (KT2)	Sao chép - dán kết quả AI mà không đọc lại, đối chiếu hay chỉnh sửa	Lỗi ảo giác trong bài nộp; mất năng lực tư duy phản biện	70% SV Y - Dược Huế (n=1.101) nhận thức AI là công cụ hỗ trợ nhưng hành vi thực tế mâu thuẫn (Nguyễn Thị Phương Thảo và cộng sự, 2024)
3. Nhận thức đạo đức mờ nhạt (KT3)	Không biết khi nào cần công bố; không rõ ranh giới gian lận và rủi ro dữ liệu	Hành vi vùng xám; bất bình đẳng giữa các SV	89,3% SV Việt Nam dùng AI nhưng chỉ ~25% được đào tạo chính thức; thiếu hướng dẫn thể chế (ASEAN Foundation và Google.org, 2026; UNESCO, 2025)

(Nguồn: Tổng hợp của tác giả từ quan sát giảng dạy và đối chiếu các nghiên cứu trước)

3.2. Đề xuất Khung hành động “3T”

Dựa trên ba khoảng trống và định hướng bởi các khung lý thuyết đã xem xét (UNESCO, 2021, 2024; Allen và Kendeou, 2024; Balalle và Pannilage, 2025), bài báo đề xuất Khung “3T” như công cụ định hướng thực hành cho SV đại học Việt Nam. Khung được tổ chức theo ba tầng lũy tiến - mỗi tầng giải quyết trực tiếp một khoảng trống hành vi cụ thể và xây dựng trên tầng trước đó. Sự tương ứng một - một giữa khoảng trống và tầng khung là điểm cốt lõi: khoảng trống kỹ năng (KT1) dẫn đến T1 - Thành thạo; khoảng trống liên chính học thuật (một phần KT2 và KT3) dẫn đến T2 - Trung thực; khoảng trống ý thức đạo đức và trách nhiệm dài hạn (KT2 và KT3) dẫn đến T3 - Trách nhiệm. Tên viết tắt “3T” (**Thành thạo - Trung thực - Trách nhiệm**) được bản địa hóa cho bối cảnh thể chế và văn hóa học thuật Việt Nam, chuyển các nguyên tắc quốc tế thành chỉ dẫn thực hành gắn gũi với SV. Khung “3T” được tổng hợp cụ thể trong bảng 3.

3.2.1. T1 - Thành thạo (Mastery): Sử dụng AI đúng cách và hiệu quả

Tầng đầu tiên giải quyết khoảng trống kỹ năng (KT1), tương ứng với thành phần Kiến thức và Đánh giá của Khung ED-AI Lit (Allen và Kendeou, 2024). Hai kỹ năng cốt lõi:

Kỹ năng viết prompt hiệu quả theo mô hình 5 yếu tố (Nhân vật - Mục tiêu - Bối cảnh - Đối tượng - Giới hạn): xác định rõ vai trò AI cần đảm nhận, mục tiêu nhiệm vụ, ngữ cảnh, đối tượng tiếp nhận và ràng buộc đầu ra. Kỹ năng này đặc biệt cần thiết và hội tụ với bằng chứng cho thấy SV Việt Nam chủ yếu dùng AI ở mức tìm kiếm thông tin đơn giản (Đặng Văn Em và cộng sự, 2024).

Kỹ năng kiểm chứng đầu ra AI gồm ít nhất ba bước: (1) đọc toàn bộ nội dung, (2) đối chiếu thông tin quan trọng với nguồn độc lập và (3) điều chỉnh cho phù hợp yêu cầu cụ thể. Đây là ứng dụng trực tiếp của nguyên tắc Human-in-the-Loop và thành phần Đánh giá trong Khung ED-AI Lit (Allen và Kendeou, 2024).

3.2.2. T2 - Trung thực (Integrity): Minh bạch và trung thực học thuật

Tầng thứ hai giải quyết khoảng trống nhận thức đạo đức (KT3). Bài báo đề xuất ba mức công bố dựa trên mức độ đóng góp của AI vào sản phẩm cuối:

Mức 1 - Không cần công bố: AI chỉ đóng vai trò công cụ tham khảo (tìm ý tưởng ban đầu, kiểm tra chính tả, dịch vai thuật ngữ); nội dung chính là công trình của SV.

Mức 2 - Khuyến khích công bố: AI hỗ trợ cấu trúc, gợi ý ý tưởng hoặc cải thiện ngôn ngữ trong các đoạn SV đã viết - AI cộng tác đáng kể nhưng không tạo ra nội dung chính.

Mức 3 - Bắt buộc công bố: AI tạo ra phần đáng kể hoặc toàn bộ nội dung, dù SV đã chỉnh sửa. Không công bố ở mức này là vi phạm học thuật.

Do tỉ lệ không công bố cao xuất phát chủ yếu từ hiểu lầm (Perkins, 2023), bài báo đề xuất tiêu chuẩn tự giám sát thực tiễn: “*Nếu tôi cảm thấy khó chịu khi nói với giảng viên rằng tôi đã dùng AI để làm việc này, sự khó chịu đó là tín hiệu tôi không nên làm vậy*”. Nguyên tắc này phù hợp với thành phần Quyền tự chủ của Khung ED-AI Lit (Allen và Kendeou, 2024).

3.2.3. T3 - Trách nhiệm (Responsibility): Con người là trung tâm

Tầng thứ ba là tư duy bao trùm toàn bộ khung, giải quyết khoảng trống về ý thức trách nhiệm (KT2, KT3), bao gồm ba chiều:

Tự trách nhiệm: Năng lực phân tích, phản biện và sáng tạo là tài sản nghề nghiệp không thể ủy thác cho AI. Nguyễn Thị Lê Quỳnh và cộng sự (2024) ghi nhận SV Việt Nam tự lo ngại về “sự phụ thuộc quá mức và mất khả năng tìm kiếm thông tin độc lập”. Khi kinh tế số ASEAN dự kiến mở rộng từ 300 tỉ lên 1.000 tỉ USD vào 2030 (ASEAN Foundation và Google.org, 2026), nhà tuyển dụng sẽ tìm kiếm SV tốt nghiệp biết phán xét khi làm việc cùng AI - không phải những người để AI thay thế.

Trách nhiệm bảo mật dữ liệu: UNESCO (2021) xác định quyền riêng tư và bảo vệ dữ liệu là một trong mười nguyên tắc đạo đức AI cốt lõi. Mặc dù Việt Nam đã có Nghị định số 13/2023/NĐ-CP về bảo vệ dữ liệu cá nhân (có hiệu lực từ 01/7/2023), nhận thức của SV về các quy định này trong bối cảnh sử dụng công cụ AI công cộng (phần lớn do doanh nghiệp nước ngoài vận hành) vẫn còn hạn chế. SV cần ý thức rõ: việc nhập thông tin cá nhân nhạy cảm, dữ liệu của bên thứ ba hay thông tin nội bộ của tổ chức vào các công cụ AI công cộng có thể vi phạm các quy định này.

Trách nhiệm trong tư duy dài hạn: UNESCO (2021) nhấn mạnh con người không bao giờ được tách rời trách nhiệm của mình khỏi quá trình tạo ra sản phẩm. Với SV: người kí tên trên bài nộp phải hoàn toàn chịu trách nhiệm về nội dung đó, bất kể cách thức tạo ra. Tinh thần này đồng nhất với Nghị quyết số 71-NQ/TW (Ban Chấp hành Trung ương, 2025): giáo dục Việt Nam hướng đến tích hợp giá trị đạo đức xuyên suốt các chương trình học.

Bảng 3. Khung hành động 3T - Thành thạo, Trung thực, Trách nhiệm (Nguồn: Đề xuất của tác giả)

Tầng	Trọng tâm	Cơ sở lý thuyết	Hành động thực tiễn
T1 - Thành thạo	Viết lệnh hiệu quả và kiểm chứng kết quả AI	Allen và Kendeou (2024): Kiến thức + Đánh giá; UNESCO (2024): Khung Năng lực AI	Áp dụng mô hình 5 yếu tố (Nhân vật - Mục tiêu - Bối cảnh - Đối tượng - Giới hạn); kiểm chứng 3 bước trước khi nộp
T2 - Trung thực	Minh bạch khi dùng AI; công bố đúng mức độ sử dụng	Chan (2024): AI-giarism; Perkins (2023); Gonsalves (2024); Allen và Kendeou (2024): Quyền tự chủ	Xác định mức công bố (1-3); áp dụng tiêu chuẩn tự giám sát đạo đức
T3 - Trách nhiệm	Con người làm chủ; bảo mật dữ liệu; phát triển năng lực dài hạn	UNESCO (2021, 2025); Nghị quyết số 71-NQ/TW (Ban Chấp hành Trung ương, 2025)	Không nhập dữ liệu nhạy cảm vào AI công cộng; phát triển kỹ năng song song với dùng AI

Khung 3T có ba điểm khác biệt so với các khung quốc tế hiện có. Thứ nhất, đây là khung tích hợp đồng thời ba chiều: năng lực AI (T1), liêm chính học thuật (T2) và trách nhiệm dữ liệu - dài hạn (T3) - thay vì chỉ tập trung vào một chiều như các khung trước. Thứ hai, khung được xây dựng từ dữ liệu quan sát thực địa tại Việt Nam, không phải vận dụng trực tiếp các mô hình phương Tây vào bối cảnh giáo dục Việt Nam. Thứ ba, cơ chế tự giám sát đạo đức đơn giản (“nếu tôi cảm thấy khó chịu...”) cung cấp một ngôn ngữ hành động dễ hiểu, dễ áp dụng để SV tự điều chỉnh hành vi - điều mà các khung học thuật trừu tượng thường thiếu.

3.3. Hàm ý chính sách cho các cơ sở giáo dục đại học

Ba khoảng trống hành vi không chỉ là vấn đề cá nhân SV mà phản ánh khoảng trống thể chế. Ba khuyến nghị chính sách được đề xuất, bám trực tiếp vào phát hiện của nghiên cứu:

Ban hành chính sách AI rõ ràng ở cấp cơ sở đào tạo: Khoảng trống nhận thức đạo đức (KT3) cho thấy vấn đề không phải ở ý định mà ở thiếu hướng dẫn. Các trường cần ban hành quy định rõ ràng - phân biệt các mức sử dụng AI được phép, được khuyến khích và bị cấm theo từng loại bài tập như bộ quy tắc ứng xử với AI - để chuyển hóa Quyết định số 1705/QĐ-TTg (Thủ tướng Chính phủ, 2024) và Nghị quyết số 71-NQ/TW (Ban Chấp hành Trung ương, 2025) thành hành động cụ thể ở cấp khoa và môn học.

Tích hợp năng lực AI vào chương trình đào tạo: Khoảng trống kỹ năng (KT1) cho thấy kỹ năng viết prompt và kiểm chứng đầu ra AI cần được đưa vào chuẩn đầu ra chương trình - không như một môn học riêng mà tích hợp xuyên suốt, dựa trên Khung năng lực số (Bộ GD-ĐT, 2025) và Khung năng lực AI của UNESCO (2024).

Trang bị năng lực AI cho giảng viên: Dữ liệu ASEAN Foundation và Google.org (2026) cho thấy 81,9% giảng viên đã dùng AI nhưng khoảng cách giữa sử dụng cá nhân và hướng dẫn có đạo đức vẫn còn rộng. Các chương trình bồi dưỡng cần bao gồm: hiểu rủi ro ảo giác AI, thiết kế đánh giá phù hợp với môi trường có AI và xây dựng văn hóa liêm chính học thuật.

4. Kết luận và bình luận

AI đã làm thay đổi căn bản cách học của SV và xu hướng này là không thể đảo ngược. Vấn đề đặt ra không còn là có sử dụng AI hay không, mà là sử dụng như thế nào một cách có đạo đức và trách nhiệm. Ba khoảng trống hành vi được ghi nhận - (1) kỹ năng đặt lệnh hạn chế, (2) nộp bài AI chưa kiểm chứng và (3) nhận thức đạo đức mờ nhạt - có sự hội tụ với các nghiên cứu trước tại Việt Nam, gợi mở xu hướng có thể phổ biến trong giáo dục đại học Việt Nam, dù cần kiểm chứng thêm ở quy mô rộng hơn. Khung 3T không phải bộ quy tắc cứng nhắc mà là định hướng dẫn dắt: AI là công cụ mạnh mẽ và giá trị của SV nằm ở chỗ sử dụng nó có chủ đích, minh bạch và có trách nhiệm. Giá trị cốt lõi của Khung 3T không nằm ở việc thay thế các quy định học thuật, mà ở chỗ tạo ra một ngôn ngữ hành động gần gũi, giúp SV tự điều chỉnh và phát triển bản sắc học thuật có trách nhiệm trong thời đại AI. Tuy nhiên, nghiên cứu có những hạn chế cần thừa nhận: thiết kế quan sát định tính không cho phép khái quát hóa thống kê; dữ liệu chưa được xác nhận bằng khảo sát định lượng đại diện toàn quốc và Khung 3T chưa được kiểm chứng thực nghiệm về hiệu quả. Hai hướng nghiên cứu tương lai được đề xuất: (1) khảo sát định lượng quy mô lớn về thái độ, hành vi và nhận thức đạo đức AI của SV đại học Việt Nam; (2) nghiên cứu thực nghiệm đánh giá hiệu quả tập huấn năng lực AI được thiết kế theo Khung 3T trong bối cảnh lớp học Việt Nam.

Tuyên bố về vai trò của các tác giả: Nguyễn Duy Hải: Xây dựng ý tưởng nghiên cứu, Thiết kế phương pháp nghiên cứu, Thực hiện nghiên cứu và thu thập dữ liệu, Phân tích dữ liệu, Viết bản thảo ban đầu, Trục quan hóa dữ liệu, Chỉnh sửa và phản biện bản thảo. Tô Hồng Nam: Cung cấp tài nguyên, Thẩm định và kiểm chứng kết quả, Quản lý và xử lý dữ liệu, Giám sát và hướng dẫn, Quản lý và điều phối.

Tuyên bố về GenAI và Quyền tác giả: Trong quá trình chuẩn bị bản thảo này, các tác giả đã sử dụng NotebookLM Plus để tổng hợp các nghiên cứu trước đó và phân tích các khoảng trống trong kết quả nghiên cứu, Grammarly Pro để kiểm tra dịch thuật tiếng Anh ở phần Abstract, ChatGPT Plus để kiểm tra logic của tài liệu tham khảo và định dạng theo chuẩn APA 7th.

Tuyên bố về xung đột lợi ích: Các tác giả tuyên bố không có xung đột lợi ích trong nghiên cứu này.

Thông tin tài trợ: Các tác giả không nhận tài trợ từ các dự án hoặc đề tài nghiên cứu.

Tài liệu tham khảo

- Allen, L. K., & Kendeou, P. (2024). ED-AI Lit: An interdisciplinary framework for AI literacy in education. *Policy Futures in Education, 11*(1), 3-10. <https://doi.org/10.1177/23727322231220339>
- ASEAN Foundation & Google.org. (2026). *AI Ready ASEAN research: First look findings* [Conference presentation]. AI Ready ASEAN: 3rd Regional Policy Convening, Manila, Philippines. https://aseanfoundation.org/programme/ai_ready_asean
- Balalle, H., & Pannilage, S. (2025). Reassessing academic integrity in the age of AI: A systematic literature review on AI and academic integrity. *Social Sciences & Humanities Open, 11*, Article 101299. <https://doi.org/10.1016/j.ssaho.2025.101299>
- Ban Chấp hành Trung ương (2025). *Nghị quyết số 71-NQ/TW ngày 22/8/2025 của Bộ Chính trị về đột phá phát triển giáo dục và đào tạo.*
- Bộ GD-ĐT (2025). *Thông tư số 02/2025/TT-BGDĐT ngày 24/01/2025 quy định Khung năng lực số cho người học.*
- Chan, C. (2023). *Is AI changing the rules of academic misconduct? An in-depth look at students' perceptions of "AI-giarism"*. arXiv:2306.03358. <https://doi.org/10.48550/arXiv.2306.03358>
- Chan, C. (2024). Students' perceptions of "AI-giarism": Investigating changes in understandings of academic misconduct. *Education and Information Technologies, 30*, 8087-8108. <https://doi.org/10.1007/s10639-024-13151-7>
- Cui, V., Trang, T. T. N., Bird, J. J., Van Hien, N., Nam, N. H., Le, Q. V., Linh, D. T., Tran, T. T. L., Wheatcroft, L., Rahman, M. A., & Brown, D. J. (2026). Diversifying understandings of AI literacy: A Global South

- perspective from Vietnam education. In E. Hart, T. Horvath, Z. Tan, & S. Thomson (Eds.), *Advances in Computational Intelligence Systems* (pp. 170-181). Springer. https://doi.org/10.1007/978-3-032-07938-1_15
- Đặng Văn Em, Nguyễn Đình Loan Phương, Nguyễn Thị Hào (2024). Thực trạng ứng dụng ChatGPT trong việc học tập, nghiên cứu của sinh viên Đại học Quốc gia Thành phố Hồ Chí Minh. *Tạp chí Giáo dục*, 24(1), 36-41. <https://tcgd.tapchigiaoduc.edu.vn/index.php/tapchi/article/view/1212>
- Gonsalves, C. (2024). Addressing student non-compliance in AI use declarations: Implications for academic integrity. *Assessment & Evaluation in Higher Education*, 50(4), 592-606. <https://doi.org/10.1080/02602938.2024.2415654>
- Nguyễn Thị Lê Quỳnh, Nguyễn Thị Ngọc Anh, Đặng Thị Quỳnh Hoa, Trương Thị Quỳnh (2024). Nhận thức về việc sử dụng ChatGPT làm công cụ hỗ trợ học tập và nghiên cứu của sinh viên Trường Đại học Ngoại ngữ - Đại học Đà Nẵng. *Tạp chí Khoa học và Công nghệ, Đại học Đà Nẵng*, 22(5B), 64-68. <https://jst-ud.vn/jst-ud/article/view/9080>
- Nguyễn Thị Phương Thảo, Bùi Mạnh Hùng, Nguyễn Thị Anh Phương, Nguyễn Huyền Trâm, Trương Mai Vinh Thoi, Nguyễn Sinh Nhật, Phạm Thị Thu Hà (2024). Thực trạng sử dụng ChatGPT của sinh viên Y - Dược Huế và một số yếu tố liên quan. *Tạp chí Khoa học và Công nghệ, Đại học Thái Nguyên*, 229(03), 403-410. <https://doi.org/10.34238/tnu-jst.9788>
- Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B. P. T. (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies*, 28, 4221-4241. <https://doi.org/10.1007/s10639-022-11316-w>
- Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching & Learning Practice*, 20(2), Article 7. <https://doi.org/10.53761/1.20.02.07>
- Phung, T. T. T., Tran, H. H. G., Dao, T. N. T., & Nguyen, P. T. (2025). Integrating AI into language learning: Insights from English majors at a university in Vietnam. *International Journal of Social Science and Human Research*, 8(4). <https://doi.org/10.47191/ijsshr/v8-i4-57>
- Thủ tướng Chính phủ (2024). *Quyết định số 1705/QĐ-TTg ngày 31/12/2024 phê duyệt Chiến lược phát triển giáo dục đến năm 2030, tầm nhìn đến năm 2045*.
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. <https://unesdoc.unesco.org/ark:/48223/pf0000380455>
- UNESCO. (2024). *AI competency framework for students*. <https://unesdoc.unesco.org/ark:/48223/pf0000391105>
- UNESCO. (2025). *Viet Nam: Artificial intelligence readiness assessment report*. <https://articles.unesco.org/sites/default/files/medias/fichiers/2025/10/Viet%20Nam%20Artificial%20Intelligence%20Readiness%20Assessment%20Report.pdf>